Technical Report ARFSD-TR-90007

METHODS FOR IDENTIFYING OBJECT CLASS, TYPE, AND ORIENTATION IN THE PRESENCE OF UNCERTAINTY

Russell W. Taylor
IBM Thomas J. Watson Research Center

Anthony P. Reeves
Cornell University

Frank P. Kuhl
ARDEC
Project Engineer

August 1990

U.S. ARMY ARMAMENT RESEARCH, DEVELOPMENT AND ENGINEERING CENTER

Armament Engineering Directorate

Picatinny Arsenal, New Jersey

US ARMY ARMAMENT MUNITIONS & CHEMICAL COMMAND ARMAMENT RDE CENTER

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER Technical Report ARFSD-TR-90007

5. MONITORING ORGANIZATION REPORT NUMBER)

6a. NAME OF PERFORMING ORGANIZATION ARDEC, FSAC

6b. OFFICE SYMBOL SMCAR-FSF-RC

7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (CITY, STATE, AND Z'P CODE) Picatinny Arsenal, NJ 07806-5000 (cont)

7b. ADDRESS (CITY, STATE, AND ZIP CODE)

8a. NAME OF FUNDING/SPONSORING ORGANIZATION ARDEC, IMD STINFO Br.

8b. OFFICE SYMBOL SMCAR-IMI-I

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (CITY, STATE, AND ZIP CODE) Picatinny Arsenal, NJ 07806-5000

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (INCLUDE SECURITY CLASSIFICATION)

METHODS FOR IDENTIFYING OBJECT CLASS, TYPE, AND ORIENTATION IN THE PRESENCE OF UNCERTAINTY

12. PERSONAL AUTHOR(S)

R W Taylor , IBM Thomas J. Watson Research Center; A.P. Reeves, Cornell University, and F.P. Kuhl, ARDEC

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (YEAR, MONTH, DAY)	15. PAGE COUNT
	FROM ________ TO ________	August 1990	39

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (CONTINUE ON REVERSE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

Model based recognition systems
Likelihood of error analysis
Fourier descriptors
Moments descriptors

19. ABSTRACT (CONTINUE ON REVERSE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

Techniques are presented for identifying unoccluded three-dimensional objects from arbitrary viewing angles in the framework of a model-based feature vector classification scheme. Fourier descriptors and moments are used for feature vector generation from contour imagery and silhouette and/or range imagery, respectively. A class of objects, airplanes, is defined with six distinct example types in our test data set. An additional data set of four objects from this class is also defined. A method for generating an exhaustive set of library views and worst case test views has been developed using a polyhedral approximation to a sphere. Based on matching to this library, object class membership,type, and orientation are determined. An approach called classification quality assessment (CQA) is applied to this recognition paradigm to both assess and deal with uncertainty. This is a two level process the first rejects objects that are not members of a known class and therefore not contained in the model database, while the second identifies the likelihood of error for classification of known object type and/or orientation (within class errors).Both use simple measures that were generated solely from the system's a priori knowledge. Experiments, using (cont)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT

☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL I. HAZNEDARI

22b. TELEPHONE (INCLUDE AREA CODE) (201) 724-3316

22c. OFFICE SYMBOL SMCAR-IMI-I

DD FORM 1473, 84 MAR

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

(KR) (

Contents

Figures

INTRODUCTION

A major computer vision task for robotics and other applications is the identification of an unknown object as a member of a set known objects. The possibility that an unknown might not belong to the set of known objects is a rarely considered type of uncertainty. The methods presented here accomplish object type identification and orientation estimation, as well as, considering several forms of uncertainty by developing models for them and rapid assessment techniques.

This report presents a unified set of techniques for the analysis of unoccluded views of unknown objects. The original image data may be of any type; silhouette, contour, and range data are used as examples . Three different types of object descriptors are presented and contrasted. Their performance is analyzed, in the presence of noise, for the basic task of estimating object orientation and/or position. To go beyond the usual model matching tests, objects are added to the test set that are not in the known object database. Techniques are then developed which reject these objects, avoiding misclassifications. Methods to assess the reliability of classifications of known objects are also developed to reduce unanticipated errors. All these techniques are framed in an intergrated architecture that is generally applicable.

The three dimensional model database for an object class is a library of feature vectors, a set for each object type instance, generated from a uniform sampling of all possible viewing angles. Objects are assumed to be rigid bodies viewed without occlusion. Self-occlusion is allowed and is handled by the library sampling technique. Unknowns in any orientation may be submitted for classification. Position is also unconstrained provided there is not a significant change in perspective distortion between model and unknown images. This implies an approximately fixed distance between the sensor and the object or sufficient distance to approximate an orthographic projection.

To produce a classified feature vector representation, an image is segmented to separate the object region from the background. Then a normalized feature vector is generated to efficiently represent the object view; feature vectors based on moments and Fourier descriptors are considered here. For both of these schemes, a fixed number of feature elements are generated for each object. This representation is then normalized with respect to location, orientation (specifically, rotation of the object in the x-y lane), and size The feature vectors each represent a single image of a three-dimensional object, viewed from a given angle. The original image data can be either silhouette, silhouette contour, or 2 1/2- dimensional range data. The term 2 1/2 dimensional indicates that 3-dimensional information is available only for the visible surfaces in the image to differentiate from tomographic (CAT) and computer aided design (CAD) data that contain full three-dimensional data for an object. Object identification is achieved by finding the best match between an unknown object feature vector and the model library. The attributes of the best match model's are then conferred on the unknown.

Classification of these feature vectors is an interesting pattern recognition problem. The continuum of views of a three-dimensional object, from either a constant or normalized distance, from a continuous surface in feature hyperspace rather than a cluster of points for a given class as is generally considered in classical pattern recognition problems. Furthermore, the differences between objects, particularly within a class, are frequently less than the variations over different views for just a single object. A nearest-neighbor classification rule is used to deal with this problem. Knowledge of the topology of feature space is used to assess the reliability of the nearest-neighbor decision.

For this recognition approach to function in an unconstrained environment, it must be assumed that the objects other than those of just the classes (and types) that are known to the system may be submitted for identification. In this situation, the classification process only needs to identify a subset of the full set of objects it might encounter. Those items not in the subset may simply be grouped together; the total number of classes is then the size of the subset plus one. The extra class encompasses all remaining objects not in the subset. In this environment, it is possible to define a post-classification analysis technique to assess classification reliability.

This is accomplished by classification quality assessment (CQA) which examines a given classification of an unknown object in relation to the available known choices and decides whether the unknown is likely to have come from the set of known classes. This is the significant difference of this approach from the work of others; CQA offers an explicit method for selecting a finite set of known objects from a potentially infinite set of unknowns. This is a post-classification test that does not increase the order of the computational complexity from that of the underlying system for identifying the finite set of objects.

CQA does nothing to speedup or simplify the basic system, other than allowing it to contain potentially fewer models. Most related research has been aimed at achieving faster searching or more effective data representations that will allow larger databases of objects. This is not the goal here; these approaches directly address the fact that the known search space is inevitably bounded by the assumption of a finite set of known objects. These methods make explicit the boundaries of search space, but do not expand it.

A second level of CQA analysis is also defined. This rejects the classification of an unknown that comes from the set of known classes, but is likely to result in selecting the wrong known object type or orientation as the identification choice. In essence , this stage attempts to recognize errors in classifying objects from the set of known object classes before they occur.

This set of techniques is intended as both a testbed for feature vector description methods and object identification tasks, and as a way to achieve generality in the presence of uncertainty. The latter was achieved without significantly increasing the complexity over the basic classification method. Most of the information needed for CQA is precomputed from the known data, and very little additional work is needed at classification time.

In previous work (ref 1), two dimensional silhouette data feature vectors based on standard moments (ref 2), moment invariants (ref 3), normalized Fourier descriptors (refs 1 and 4), and three dimensional standard moments (ref 5), have been compared. Results indicated that the performance of standard moments was far better than moment invariants and slightly better than Fourier descriptors for the given task. The use of the descriptors for determining orientation was first demonstrated in reference 6. The use of three dimensional information is critical for disambiguating orientation possibilities for objects with global symmetries. This report will also consider the application of CQA methods (refs 7 and 8) to orientation estimation.

Important aspects of this work include the development of a quantitative method for comparing different shape analysis schemes, generation of realistic synthetic data, effective feature vector descriptors, and ways to dynamically assess classification accuracy. Additionally, all these considerations are brought to bear on the problem of estimating the orientation of objects that belong to a class of similar object types.

STRUCTURE OF THE IDENTIFICATION SYSTEM

The techniques developed are collected together in the object detection and classification system shown below in figure 1. The system may be considered in two sections; the precomputed part which is essentially the training phase, and then the run time portion which is used for testing. For training, synthetic range images are generated from parameterized three-dimensional shape descriptors. For real images, a camera and automatic positioning system could be used. Normalized feature vectors computed from these images are stored in a library for all classes. For this particular classification task, feature vector balancing has been found to improve the performance of the system (ref 1). This involves computing the standard deviation of each feature vector element over the entire library. Individual feature vector elements are then balanced by dividing their associated standard deviation. Balanced feature vectors are then analyzed for CQA considerations and stored in the final overall library database.

For controlled experiments, synthetic data are used for testing. It is also possible to use real imagery for testing, as documented in previous two-dimensional experiments (ref 9). Synthetic test images are generated from the same parameterized descriptions used to generate the library data but with different viewing angles and resolution.

Noise is then added to these ideal images. A range noise model was developed to add meaningful noise corruption to the range image data.

Once an image is obtained, an umage processing stage is performed. This stage is responsible for noise filtering and object segmentation . A variety of classic image processing techniques can be used to reduce the effects of noise given, typically, a sensor-based noise model. For the moment-based feature vectors, no noise filtering was done for the current experiment's synthetic noisy range data. For Fourier descriptors, 3x3 mean filtering was performed to guarantee that a single closed contour could be extracted. The range images were segmented by simple depth thresholding.

Following the preprocessing stage, a normalized feature vector is generated for the segmented region in the test image. This vector is balanced using the same parameters used to balance the library vectors. The object type identification is then made by finding the best match between the library of feature vectors and the test image feature vector. The orientation of the object is also obtained from the selected library entry. The quality of this decision is then examined by the CQA process to determine its validity.

MODEL DATABASE GENERATION

Icosahedron Based Sampling

To reasonably assess the success of any classification scheme, an unbiased and recreatable set of test data is required. Prior to the work accomplished in reference 5, most analysis of object recognition success from different viewing angles was performed for just a set of random sample views that might, or might not, encompass the degenerate points. In order to fully evaluate the object identification effectiveness of a global feature representation and matching metric, an exhaustive, worst case technique was developed. This was tested with the different feature vector representations.

Given a three degree of freedom viewing angle problem, the ideal test data set would consist of all possible viewpoints. Based on initial assumption and range normalization performed, this can be reduced to all possible viewpoints on a spherical surface where the object is positioned at the sphere's center. This is still an impossible data set to generate; the next best choice is to equally partition the sphere's surface into a grid of viewpoints. These viewing locations would be close enough to insure that no problematic views were omitted.

Such a uniform partitioning of a spherical surface is a nontrivial problem. A good approximation, however, can be made by using a polyhedron with its sides subpartitioned to achieve the desired resolution level. Ballard and Brown (ref 10) suggest the use of an icosahedron. A number of researchers have made use of this type of approach to achieve a partitioning of the Gaussian sphere into an extended Gaussian image (refs 11 and 12).

An icosahedron was used for our current experiments. This 20-sided polyhedron had each of its sides subdivided into 25 equilateral and identical triangles. The resulting polyhedron, when expanded so that all vertices touched the surface of an enclosing sphere, had 500 faces and 252 vertices. Two data sets were then defined, one composed of viewing angles corresponding to polyhedron's vertices and the other viewpoints located at the center of each side. One data set was designated as the library (known) views and the other as the unknown views. The results is a worst-case test of a given classifier, since the viewing locations for one data set are interstitially located between the viewing locations of the other set. Therefore, the unknown views are always as far as possible from the library views in geometric space The relationship between the icosahedron and the viewing locations is shown in figure 2, and an actual idea of the density of samples for the two sets of views in figure 3.

The viewing angle difference between library and unknown views of an airplane vary from 6.0 to 8.7 degrees, with an average of 7.3 degrees, using the polyhedral tessellation technique. It is assumed that the less than 3 degrees variation range is not a problem. Wheather or not the overall tessellation is fine enough is assessed by the CQA tests.

The Airplane Class

Starting with the airplane identification experiments of earlier researchers (refs 3 and 4), a systematic experimental method that will encompass a uniform subset of all possible viewpoints of a given object was defined. Additionally, we have a predefined test that will yield worst-case results for a given definition of the classification task. The only class of objects the system knows of are airplanes; the system's a priori knowledge is contained in a library of six instances (types) of the class airplane.

Each airplane is described by a short list of parameters. These parameters are a CAD-type representation of the three-dimensional surface of an airplane consisting of a set three or four sided polygons. The nonairplane objects are similarly described. Synthetic range images with specified viewing angle and resolution are created from this description by silhouette and range data rendering software. The feature vectors are than generated from these images.

Bhanu (ref 13) demonstrated that a geometric representation such as one used by the authors for aircraft, can be generated from real objects using range data. Using such a technique to generate a three dimensional model, which in turn is used to generate the model database of the authors, would maximize precomputation and minimize classification time requirements. This is one logical approach to extending our system to arbitrary real objects.

The viewing angle sets previously described were used to generate a library of normalized feature vectors; views for each airplane were generated and then collated into a library database. This library contained 500 views for each airplane; the viewing angles were obtained from the centers of the faces of the partitioned icosahedron.

Two different sets were defined. Both were at lower resolution than the library views, and one had added range noise. Each test set consisted of 252 worst-case views of the six airplanes or of the four nonairplanes (i.e., with viewing angles taken from the vertices of the partitioned icosahedron as described above).

For the trails, each test view was matched, to its closest entry in the library by means of a Euclidean distance measure in feature space. The Euclidean distance between two feature vectors, a and b, of lengths $|\vec{a}| = n$ and $|\vec{b}| = n$ is defined as:

$$D_{\varepsilon}(a,b) = \left[\sum_{i=0}^{n-1} (a_i - b_i)^2\right]^{1/2} \qquad (1)$$

The best match between an unknown and the library yields the minimum D_{ε}, notated as $D_{\varepsilon\text{-MIN}}$. This choice was then subjected to CQA considerations, and decisions was rendered.

For the experiments in this report, the full library was used for both type orientation trails. It is worth nothing that a reduced "class--not type" could be generated for orientation only experiments. In this case , only class membership (not type) would be verified, and the orientation estimated. This provides the definition of class: a set of object types that must be distinguished from each other, but that differ superficially enough that they can be geometrically registered in a meaningful way.

Range Data Noise Model

A noise model for range imagery is significantly different from a model for conventional intensity imagery, especially where discontinuities occur. For the intensity imagery, simulating focus/lens errors and sensor noise can adequately be accomplished by low pass filtering and adding (Gaussian) noise to the image (e.g., ref 1). For range imagery, the blur process is generally not meaningful, and sensor quantization and noise can produced special problems (see reference 14 for an overview of range finding techniques). For example, at an edge discontinuity in the range image, a range sensor will detect one surface or the other but not the average between the two. There is also, typically, some spatial uncertainty about the location of the edge. Approaches, such as range inference from structured light projection (ref 15) and laser time-of-flight sensors, will similarly make errors at surface discontinuities.

In the airplane experiment, the most significant edge is between the object and the background since this also affects the value of the silhouette moments. To simulate range noise realistically, the object edges in the synthetic image were perturbated by a maximum of one pixel. An object edge element which is reassigned to the background is simply removed; however, range values must be generated for background elements which are reassigned to the object. New range elements were estimated by computing the mean of the existing adjacent range values of the actual object.

A second likely source of error in the range image is the range distance measurement itself. To stimulate this form of error, Gaussian noise was added to the object range pixels. This will generally only affect the range moment values and not the silhouette moments.

Sample images from the airplane and nonairplane test set, both with and without noise, are shown in figure 4. The nonairplane test is composed of a wine bottle, a tetrahedron missing one face, an object composed of cubes, and a space shuttle.

The library views were generated at a resolution of 128x128 with depth quantiized as an integer between 0 and 127. The two test data set both had a resolution of 96x96 (x96); one was generated with noise and one without. For the noisy test set, the probability of an edge element changing from object to background (or vice versa) was set at 0.4, and the added Gaussian noise had a standard deviation of 3.0. The typical thickness of an airplane body with 96x96x96 resolution is 6 pixels. This data set is of lower quality than data that can be obtained from current dense-sampling range imaging devices [the ERIM or Technical Arts White Scanners, for example (ref 16)].

FEATURE VECTOR GENERATION AND ANALYSIS

Moment-based Feature Vector Generation and Normalization

For this work, the moments for both the range and the silhouette images of the object were computed. An image silhouette is a binary valued projection of the visible object surface function onto the x-y plane. Previous work (ref 1) has shown that the moments of the silhouette are suitable for shape classification. The best results have been obtained by using a combination of these two moment sets and by using normalization parameters of the silhouette moments to normalize the range moments. By maintaining the spatial correspondence between the two moment sets, not only are there two distinct data descriptions of the object, but also information contained in the correspondence of the two data sets. Therefore synergistic improvement in results is realized from the combination.

The conventional definition of the two-dimensional moment of order n, where n = (p+q), of a function f (x,y) is (ref 17)

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y)\, dx\, dy$$

$$p,q = 0,1,2 \qquad (2)$$

A set of momemt values may be used to represent a segment of an image. In this case, f(x,y) is the image function in the segment region. The image function is assumed to be zero outside the segment region. Transformations such as rotation, translation, and scale change can be performed in the moment domain with a small, fixed number of operations. Depending on the segment size, this can represent a substantial speedup over doing the equivalent operations in the original pixel domain. Furthermore, a truncated set of moments offers a more convenient and economical representation of the essential shape characteristics of an image segment as compared to a pixel based representation. A complete moment set (CMS) is a truncated moment set which contains all moments of order n and lower. On such a set, the operations of the scale change, rotation, and translation are closed.

The first stage of the normalization process is to compute a standard moment set for the silhouette moments (ref 2). A set of standard moments is a CMS which has been normalized with respect to scale, translation, and rotation. To define those normalizations, analogies to the moments of inertia of solid objects are drawn.

To normalize transforms are computed so that the low order transformed standard moments have the following values (ref 5):

$M_{00}=1$ (area normalized to 1)

$M_{01}= M_{10}=0$ (central moments; the centroid of the object function translated x=0, y=0)

$M_{11}=0$ (rotation normalization; the silhouette rotated to align its principal axis with a coordinate axis)

For rotation normalization, to make the rotation quadrant unique, make

$M_{20} \geq M_{02}$(the major principal axis is the x axis)

$M_{30} \geq 0$ (the projection onto the x axis has a negative skew)

improved results have been obtained by using aspect ratio normalization (ref 1). This transforms the ellipsoid of inertia of the moments to be a circle; i.e., for aspect ratio normalized moments

$$M'_{20}=M'_{02}$$

This transformation is performed after rotation normalization. The original aspect ratio of the moments, $(M_{02}/M_{20})^{0.25}$, is used in place of M_{02} as a feature vector element.

Once the silhouette moments have been normalized, the range moments are normalized with the same transform parameters. This maintains an exact correspondence between the silhouette and range moments.

An additional normalization step is required for range data, the normalization for translation in the depth dimension. Several schemes for depth normalization have been considered in previous work (ref 18), and the one documented in reference 15 is used here. The problem is that the back of a range image cannot be seen; therefore, the location of the actual center of gravity in the depth of dimension is not known. For convenience, it is assumed that the object has a flat back parallel to the image plane, and that the cross section of the occluded part of the object has the same shape as the occluding boundary (the perimeter of the silhouette). Finally, it is assumed that the normalized volume of the object is 1 (recall that the area of the normalized silhouette is also1). Any set of robust, consistent assumptions could be used; the advantage of the proposed set is that they are easily implemented in the moment domain.

The origin of the depth dimension is set to the location of the assumed back surface object. The translation of the range moment ser (R_{pq}) is achieved with

$$R'_{pq} = R_{pq} + \alpha S_{pq}$$

where (S_{pq}) is the set of silhouette moments and α is computed to set the volume to 1. For objects which are much deeper than they are wide, α will be negative; this implies that the occluded section has a negative volume. The experiments indicate that this does not cause problems. However, a slightly better performance is achieved by using a value for α which sets the volume to 3.

Fourier Descriptors Generation And Normalization

A possible set of features used to describe a contour are Fourier descriptors (FD) (refs 9,19, and 20). The method used in the experiment is fully described in reference 4. A short description of the process is included here.

Given a silhouette of an object, its contour can be extracted. If this is considered as a closed contour C lying in the complex plane, then its Fourier series can be defined. Trace it once in the counter-clockwise direction with uniform velocity v, obtaining the complex function z(t) with parameter t. Choose v so that the time T required to traverse the contour is 2π. Traversing the contour more than once yields a periodic function, which may be expanded in a convergent Fourier series. A Fourier descriptor of C is defined to be the complex Fourier series expansion of z (t).

$$z(t) = \sum_{n=-\infty}^{\infty} A(n)\, e^{jnt} \tag{3}$$

where

$$A(n) = \frac{\pi}{2} \int_{0}^{2\pi} z(t)\, e^{-jnt}\, dt \tag{4}$$

The FD depends on both C and the starting point of z(t). In practice, C is taken from a digitized image ; therefore, z(t) is not available as a continuous function. If z(k) is a uniformly sampled version of z(t) of dimension N, the discrete Fourier transform directly gives the N lowest frequency coefficients A(n).

The FD can be computed by resampling the sequence of perimeter points to span a power of 2 number of points and then computing the FFT (ref 4). Alternatively, the Fourier coefficients can be calculated using a DFT approach which uses the piece-wise linear nature of a chain code representation of the perimeter sequence to speed the calculation (ref 20).

The frequency domain operations, which affect the position, size, orientation, and starting point of the contour, follow directly from properties of the DTF. Translation is normalized by setting A(o) to zero; size is normalized by dividing all A(i) by |A (1)|. Finally, in-plane rotation and starting point position are normalized by changing the phase of the coefficients so that A(1) and A (k), the next largest coefficient, have a phase of zero.

The CQA Enhancements to the System

The CQA test is defined in two levels. When an unknown feature vector is submitted to the system, its nearest neighbor in the library of known feature vectors is found. The first CQA level examines this decision, and makes a judgment about the likelihood that this object belong to the selected class of known objects.

If an object has not come from the set of known objects, then the classification decision is disregarded.

The object is, in effect, classed as rejected by the system.
If the object passes this first CQA level, the classification decision is scrutinized at the second CQA level,. Here the object is believed to be from the set of known objects, and the CQA tries to determine how likely it is that the wrong one of these objects has been selected as the classification choice.

Both these methods can function in one of two ways. The first way is that an empirical analysis can show, based on the CQA measures, what the probability of error at each level is. This information can then be passed along with the classification decision, perhaps tempering later dependency on its validity. The second approach is that the same empirical analysis can be used to set thresholds that guarantee a specific level of certainty. The classification system then actively rejects objects that do not meet the certainty criteria.

Since the initial classification decision is made using a Euclidean distance measure, the Euclidean distance has been established as a metric to measure match quality. It is reasonable to suppose that a simple CQA assessment might be achieved by just thresholding this metric. Our CQA techniques go beyond this, trying to embody more sophisticated information about feature space, while still depending on a priori information. To show effectiveness of these methods, they are compared to simple Euclidean distance thresholding technique in the results.

Known-Class CQA

The known CQA method for determining if an unknown object comes from a known class of objects exploits the geometry of feature space. O_i (j) is defined as object i viewed from position j, and F(·) as the feature vector generating function. From this, the feature vector has been obtained for $o_i(j)$ as $\vec{v}_i(j) = F[o_i(j)]$, where $F{:}R^3 \rightarrow R^n$ given $|\vec{v}| = n$. If the n-dimension feature vectors could be generated for every possible view of an object from a fixed distance, the unnormalized feature vectors would define a continuous closed hyper-surface in feature space. This assumes that our feature vector generating function and our input image are both continuous. After normalization, there may be a small number of hyper-surface discontinuities introduced between geometrically adjacent library views (ref 4).

Any feature vector of the object will lie on this hyper-surface. A feature vector belonging to any other object will not lie on this hyper-surface, unless the degenerate case of two different objects that appear identical from view is present. This latter condition cannot be guarded against in a system that works with single views of an object.

In practice, a discrete valued feature vector generation function applied to discrete valued image function is used. Additionally, only a finite sample of the set of possible viewing angles is selected. So the feature space contains a set of points for a given object that hopefully lie close to, but probably not on, the ideal hyper-surface.

The goal of our polyhedral tessellation of physical space is to generate a fine enough, and regular enough, sampling of viewing angles to create a good approximation of the desired continuous surface in feature space. By good approximation, it is meant that any irregularities in the feature space surface are represented clearly in our sampling. A set of library samples for a known object in a three dimensional feature space and feature vectors for unknown objects of both the same and different types are shown in figure 5. An error is made by matching the unknown that does not lie on the hyper-surface shown to a library point that does.

The first stage of CQA tries to avoid the type of erroneous classification shown in figure 5. To do this, each library view has associated with it a measure of the local, same-class variability. This can be envisioned as a measure of the surface smoothness and fineness of tessellation in feature space around that library point.

The Euclidean distance, D_{ε} between an unknown feature vector and its best match library feature vector is defied as

$$D_{\varepsilon\text{-MIN}} = \min_{i\,\text{and}\,j} D_{\varepsilon}(\vec{v_i}(j), \vec{v_u}) \tag{5}$$

where i indicates object type, j viewing position, and $\vec{v_u}$ an unknown feature vector, is normalized by the library selection's variability measure. For example, if $D_{\varepsilon\text{-MIN}}$ selects object type k from view point 1, then

$$D_{CQA_1} = \frac{D_{\varepsilon\text{-MIN}}}{CQA_1(k,l)} \tag{6}$$

Empirical data are then used to assess the likelihood that the unknown feature vector is known object, base on D_{CQA1}.

Two different measures of library variability have been used. Both examine all the library viewing points in physical space neighborhood about the viewing point of interest. Therefore, n viewing points within α degrees of our viewing point of interest will be selected, and the variability statistics based on feature vectors associated with all these viewing points will be developed.

The first measure is the standard deviation of D_{ε} between the library view of interest, and the library views of the same object type in the selected neighborhood:

$$CQA_1(i,j) = \left[\frac{1}{n}\sum_{k \in nhd(j)} (D_{\varepsilon}[F(o_i(j)),F(o_i(k))] - \overline{D_{\varepsilon}})^2\right]^{1/2} \tag{7}$$

where nhd(j) is the neighborhood of j, with n viewing locations in it. The notation emphasizes that the neighborhood is in physical viewing space and not in feature space.

$\overline{D_{\varepsilon}}$ is the mean value for D_{ε} in that neighborhood. The second measure is the maximum D_{ε} between the view of interest and the neighborhood views:

$$CQA_1'(i,j) = \max_{k \in nhd(j)} D_{\varepsilon}[F(o_i(j)),F(o_i(k))] \tag{8}$$

The second method is a simple order statistic-type approximation of the first but yields superior results. Since these measures are precomputed from our a priori knowledge, the complexity of calculating the measure has no effect on classification time requirements.

Intra-class CQA

Once an unknown feature vector has passed the first level of CQA, it can be said with a computed level of certainty, that the unknown object is one from a known class. The second level of CQA now addresses the probability that our specific instance selection (type and/or orientation), from the set of known object types in that class, will be in error. Whereas, the first CQA level considers the possibility of membership in a class known to the system, in the second level of CQA this is taken as a certainty, and the quality of the specific instance identification is addressed.

Misclassification arises from several sources. For high quality data, a significant cause is insufficiently fine viewing point sample. Additional problems exist for lower quality data; noise distortion and reduced spatial resolution can make similar object indistinguishable. These last two problems are often dependent on the unknown data, rather than the known data. Consequently, it is difficult to make allowances specifically for data quality using precomputation bases on the a priori known database.

For level two CQA, the focus has been made conceptually on misclassification due primarily to insufficient sampling. An example of how this type of misclassification occurs is shown in figure 6. Here, a feature vector belonging to known type is misclassified as another type, because the library sampling is insufficient to characterize the feature space hyper-surface. The additional problems of noise and resolution are not really independent of library sampling density. An adequate amount of sampling at a given noise and resolution level may become inadequate under the burden of increased feature vector variability due to an increase in noise and/or decrease in resolution.

The initial method used for level two CQA was to normalize the Euclidean distance from an unknown feature vector to the best match library selection by a confusion factor, CQA_2. The confusion factor was the smallest of the Eiclidean distances between the selected library feature vector and all other library feature vectors for objects of different type or orientation. For example, to determine the type confusion factor for a library feature vector of airplane 1, that vector would be compared to all library feature vectors of planes 2 through 6. specifically;

$$CQA_2(i,j) = \min_{k \neq i \text{ and } l} D_{\varepsilon}(\vec{v_i}(j), \vec{v_k}(l)) \quad (9)$$

gives the formula for determining CQA_2 for type identification. The notation emphasizes that this measure is dependent on the goemetry of feature space. For orientation testing, all choices of a different orientation could be searched;

$$CQA_2(i,j) = \min_{l \neq j} D\varepsilon(\vec{v_i}(j), \vec{v_k}(l)) \quad (10)$$

but beacause of the dense sampling of views for a given type, this was not seen as a useful measure. So, for orientation CQA_2, the measure from equation 9 is used; the success is simply measured based on a different criterion.

As with the first level of CQA, the normalized Euclidean disrtance,

$$D_{CQA2} = \frac{D_{\varepsilon\text{-MIN}}}{CQA_2}$$

is then used to generate an emperical estimate of the probability of error.

EXPERIMENTAL RESULTS

The results are presented here in two sections to more clearly illustrate the behavior of different parts of the system. The first section discusses the results for type and orientation determination for the airplane test sets using the different types of feature vectors. The second section illustrates the additional effects of applying CQA analysis to the system. Results for this latter section are generated by combing the airplane and nonairplane test sets.

Type and Orientation Identification

This section presents results for Fourier descriptors, silhouette moments, and a combination of silhouette and range moments used to determone airplane type and orientation. These feature vectors are based on two dimensional contours, two dimensional slihouettes, and 2 1/2-dimensional range imagery, respectively. Type classification is correct if the correct airplane of the six possible is obtained from the best library match. The angle classification is correct if the angle of the library entry of the best match is one of the nearest neighbors in viewing space to the test shape. For our experiments, this means that the angle difference must be less than 10 degrees. However, further considerations are necessitated by the shape symmetry of the airplanes.

Because of the bilateral symmetry of airplanes, any projection of a three-dimensional view into two dimensions (e.g., a contour or silhouette) will be inherently ambiguous. That is, for any given viewpoint, there exits an associated different viewpoint from which the two dimensional projection of the object shape would be identical (possibly differing by an image plane rotation which is moved by normalization). For example, the silhouette of an airplane viewed from directly above is identical to the silhouette of that airplane viewed directly from below. Since it is impossible for the system to distinguish between these two conditions, a folded angle criteria is defined in which an angle classification is said to be correct if either of the two ambiguous viewpoints are selected by the system. In a practical environment, other cues such as direction of motion or multiple views could be used to disambiguate between the two possible angle interpretations.

Results for both the noiseless and noisy test sets for the three types of feature vectors are shown in figure 7. The moment-based feature vectors are all balanced; this was found to improve the success rate by 3 to 5% (ref 5). For Fourier descriptors, feature vector balancing was found to reduce the success slightly, especially when a large number of feature vector elements were used. This is most probably due to the emphasis of high frequency components that are highly sensitive to contour perturbation caused by noise. Consequently, all Fourier descriptor results are presented without balancing.

The following general observation were made from figure 7:

1. Little improvement for any technique in using more than 20 feature vector elements for classification.

2. Confirmation for our findings that moments are, in general, more effective than Fourier descriptors as feature vectors, and are more robust with respect to noise. For moments, range moments alone (results not shown here) produce similar type identification results to silhouette moments alone, but the two combine synergistically to produce markedly better results.

3. The majority of errors for the two dimensional image-based measures stem from the symmetry ambiguities. In most cases, the folded angle results are slightlyworse (several percent) than the type classification results. Folded angle consideration produces only a small improvement for combined silhouette and range moment feature vectors (similar behavior is seen for range moments alone).

Applying CQA

This section relates the results of applying the two levels of CQA to class, type, and orientation classification. The same three feature vector generating methods as in the previous section are used here. A single feature vector length greater than 20 elements was considered for each feature vector element type; a vector length that corresponded to a natural subset for that feature vector type was selected. The vector lengths were: Fourier descriptors (ref 23), silhouette moments (ref 24), and silhouette and range moments (ref 25). In this case, only the noisy test results are presented, and only with folded angle testing.

For CQA_1, the goal is to determine if the current unknown view is an airplane. These results where the number of nonairplanes rejected is plotted against the number of airplanes rejected for different thresholds of the confidence measure is shown in figure 8. From this graph, it is possible to select a threshold that will optimize the performance based on the relative costs of accepting a nonairplane and rejecting an airplane

The Euclidean distance is used as a baseline for all methods. In all cases, the Euclidean distance was a good confidence measure that enabled over 80% of the nonairplanes to be rejected without rejecting more than 10% of the airplanes. The performance of the Fourier descriptors was better than for silhouette moments when the threshold is set to reject only a small number of airplanes (more than 10%) are rejected. The performance of the silhouette and range moments was inferior when a small number (less than 5%) of airplanes are rejected but superior when a large number of airplanes (15%) are rejected.

The CQA_1 confidence measure is the standard deviation of variations in feature space corresponding to small view angle perturbations in local geometric space. When compared to the Euclidean distance alone, this gave very poor results for the Fourier descriptors but much better results for both moment methods and low airplane rejection thresholds. However, the Euclidean distance gave superior results for silhouette moments for a large rejection threshold.

The maximum distance to a local neighbor metric shows excellent results for the silhouette and range moments and a small improvement over Euclidean distance for silhouette moments except for high airplane rejection levels. For Fourier descriptors, this method gives slightly worse results than the Euclidean distance. From this single test, it would seem that the maxdist method is the most appropriate confidence measure to use for moment feature vectors, and the Euclidean distance is the best , by a small margin, for Fourier descriptors.

The test for CQA_2 is the minimum distance from the matched library entry to a library entry of a different type. The effect on the classification success after low confidence responses have been rejected is shown in figure 9. Low confidence responses are determined by a threshold on either the Euclidean distance to the matched library entry or the CQA_2 measure. Results are shown for both type and folded angle classification.

In all cases, the CQA_2 measure is a dramatic improvement over the Euclidean distance for type classification, especially when the reject set is large. The most marked improvement is for the silhouette and range moment feature vectors. The result of using CQA_2 in this form for angle classification is not good. A slight improvement over the Euclidean distance is noted for Fourier descriptors; however, for the moment feature vectors, the results are worse than the Euclidean distance. This is not a surprising result; the error model used for type errors is inadequate for modeling orientation errors. Future work will produce a more sophisticated and appropriate model for this phenomenon.

CONCLUSION

A systems approach to identifying objects from global feature descriptors has been presented. The importance techniques and results discussed included;

1. An exhaustive (in viewing the angle sense) worst case test set and testing procedure for arbitrary feature vector descriptors.

2. Results for orientation and type classification with and without uncertainty analysis.

3. Improvements over Euclidean distance thresholding, for the rejection of objects not included in a class of known objects, by considering local hyperspace behavior (CQA_1).

4. Improvements over Euclidean distance thresholding, for enhancing classificationtion success, by rejecting some inputs as being likely to result in errors (CQA_2).

5. Integration of the above techniques in an architecture that emphasizes precomputation and compilation of the known object database and CQA measures, to allow simple runtime analysis.

The results have been for a single experiment involving six airplanes; therefore, they may not generalize well to other applications. However, for this experiment consistent performance trends for Fourier descriptors and silhouette moments are seen, with the moments producing superior results for both type and orientation classification. Incorporating range information in the feature vector descriptors, with the silhouette and range moments, produces dramatically better results than for the two dimensional methods for type identification. A selection of classification results for type identification and orientation are summarized in tables 1 and 2, respectively.

Results for orientation classification have been inferior to the type classification results in all cases. Future work will be directed to the development of techniques for improving the orientation classification.

Table 1. Type classification results

	Fourier descriptors (23 elements)	Silhouette moments (24 elements)	Silhouette and range moments (25 elements)
Original test data	90.48	94.18	94.91
Noisy test data	84.26	93.12	94.31
CQA_2 (5% reject)	86.43	94.43	95.96
CQA_2 (10% reject)	88.31	95.81	97.35
CQA_2 (25% reject)	93.30	98.06	98.67

Table 2. Orientation (folded) classification results

	Fourier descriptors (23 elements)	Silhouette moments (24 elements)	Silhouette and range moments (25 elements)
Original test data	90.80	97.19	90.28
Noisy test data	82.41	84.72	90.60
CQA_2 (5% reject)	84.20	90.53	90.95
CQA_2 (10% reject)	84.94	91.62	91.55
CQA_2 (25% reject)	87.65	92.68	92.24

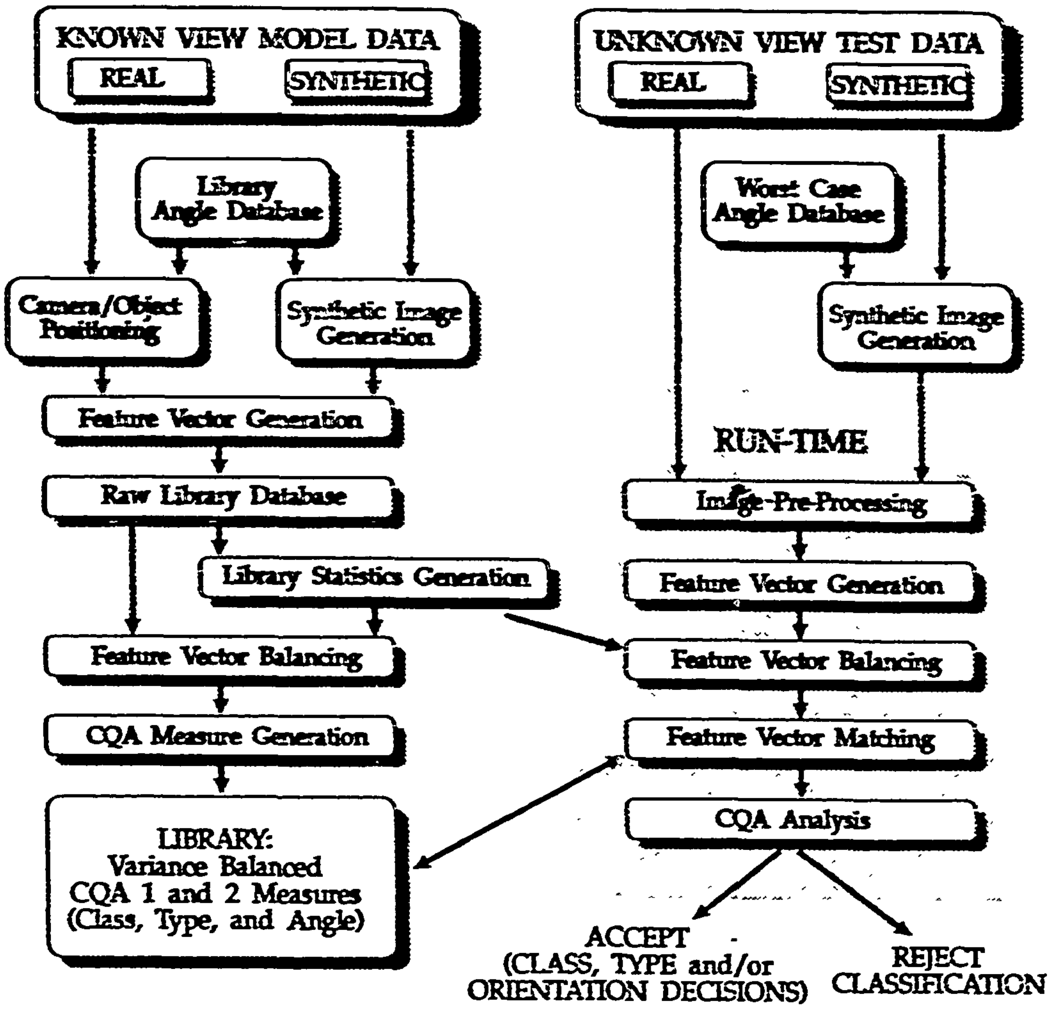
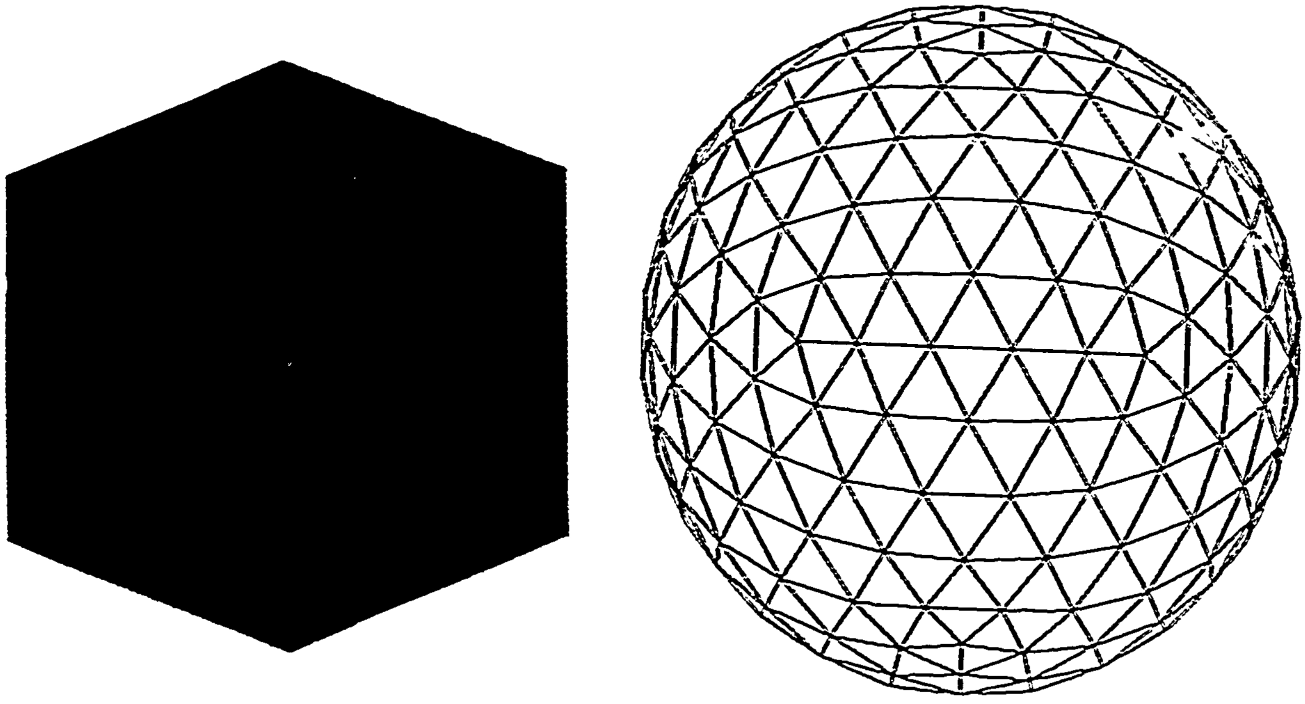


Figure 1: The overall structure of an object identification system using the techniques developed in this report. Note the emphasis on precomputation and compilation of necessary data bases using a priori knowledge.



(a)

(b)

Figure 2. An icosahedron (a) and the partitioned surfaces that define the locations of library views mapped to the surface of a sphere (b)

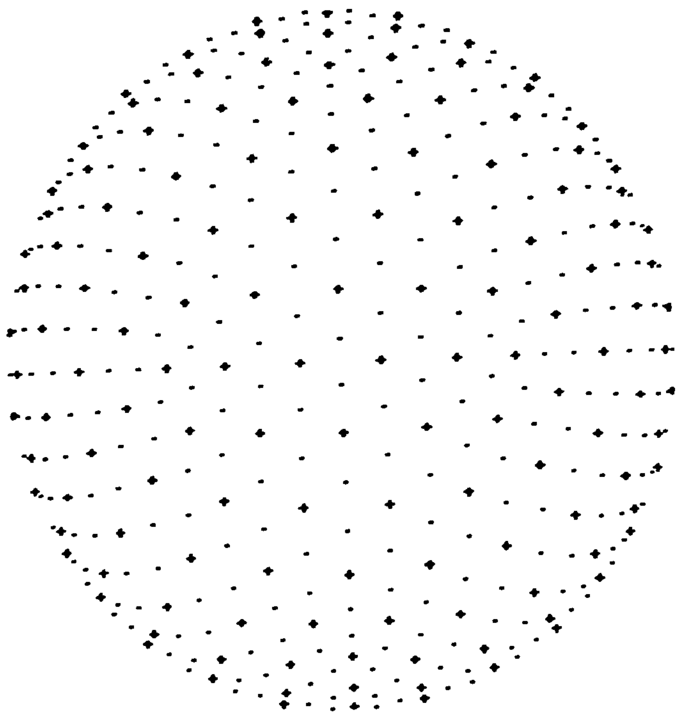
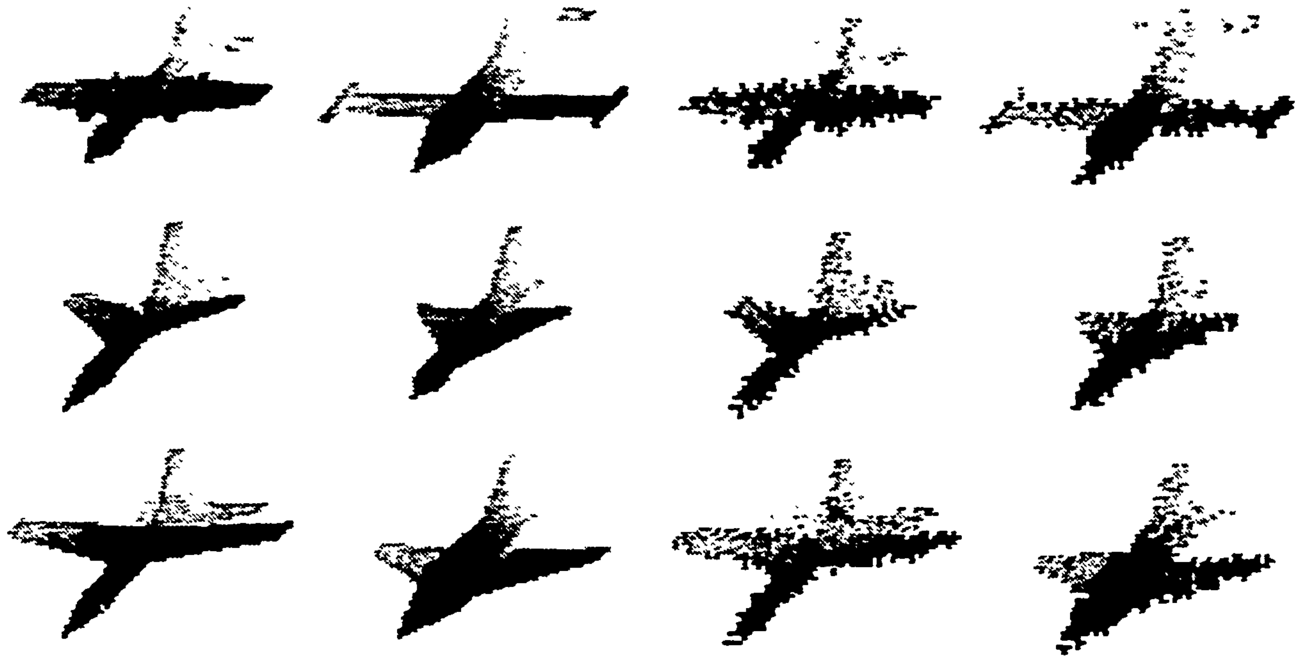


Figure 3. The viewing sphere showing (a) model library viewpoints (marked with ·) and the worst case viewpoints



Airplane object

Nonairplane objects

Figure 4. The test images for a single viewpoint shown with and without added noise (resolution of the noisy test set is 96x96x96).

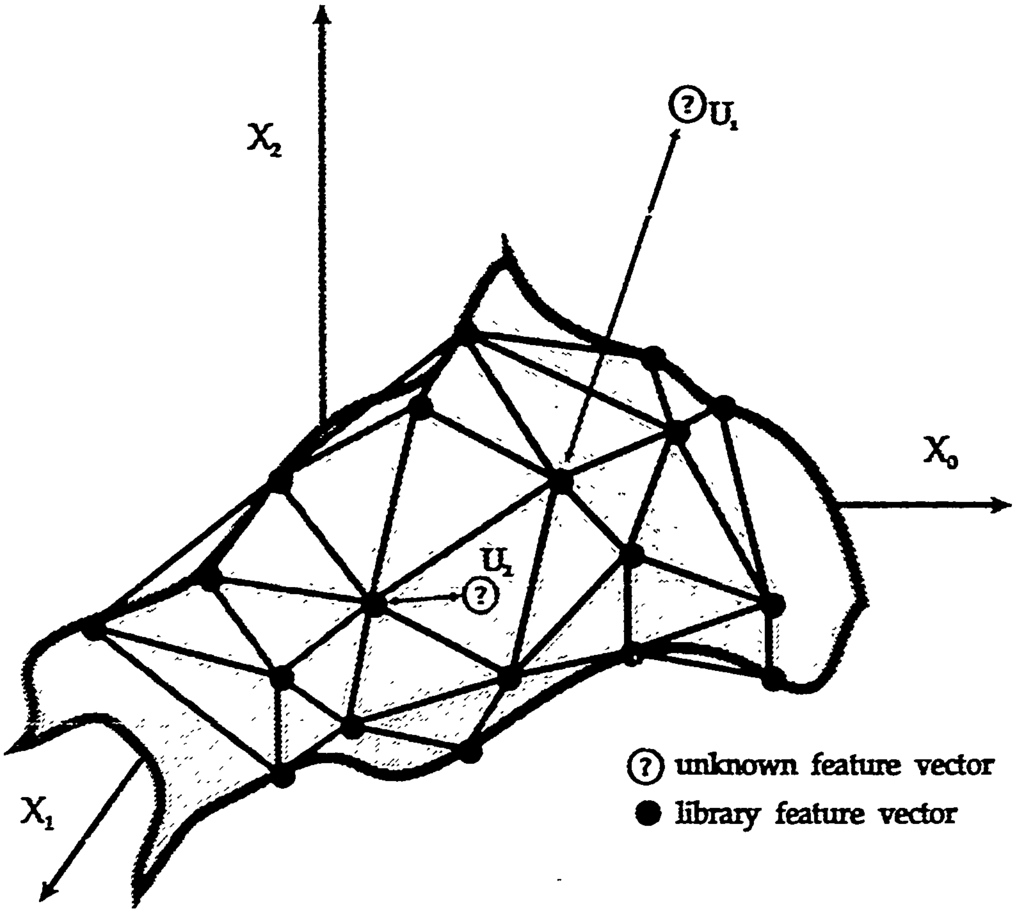


Figure 5. A three-dimensional feature space of correct and enormous identifications of unknown feature vectors (Feature space surface associated with a known object is shown in gray; known feature vectors and their nearest neighbors are marked. Unknown feature vectors are shown with their closest match known feature vectors. Unknown U1 does not lie on the surface of allowable values and is therefore incorrectly matched to the known object. U2 does not lie on the surface and is all correctly matched.)

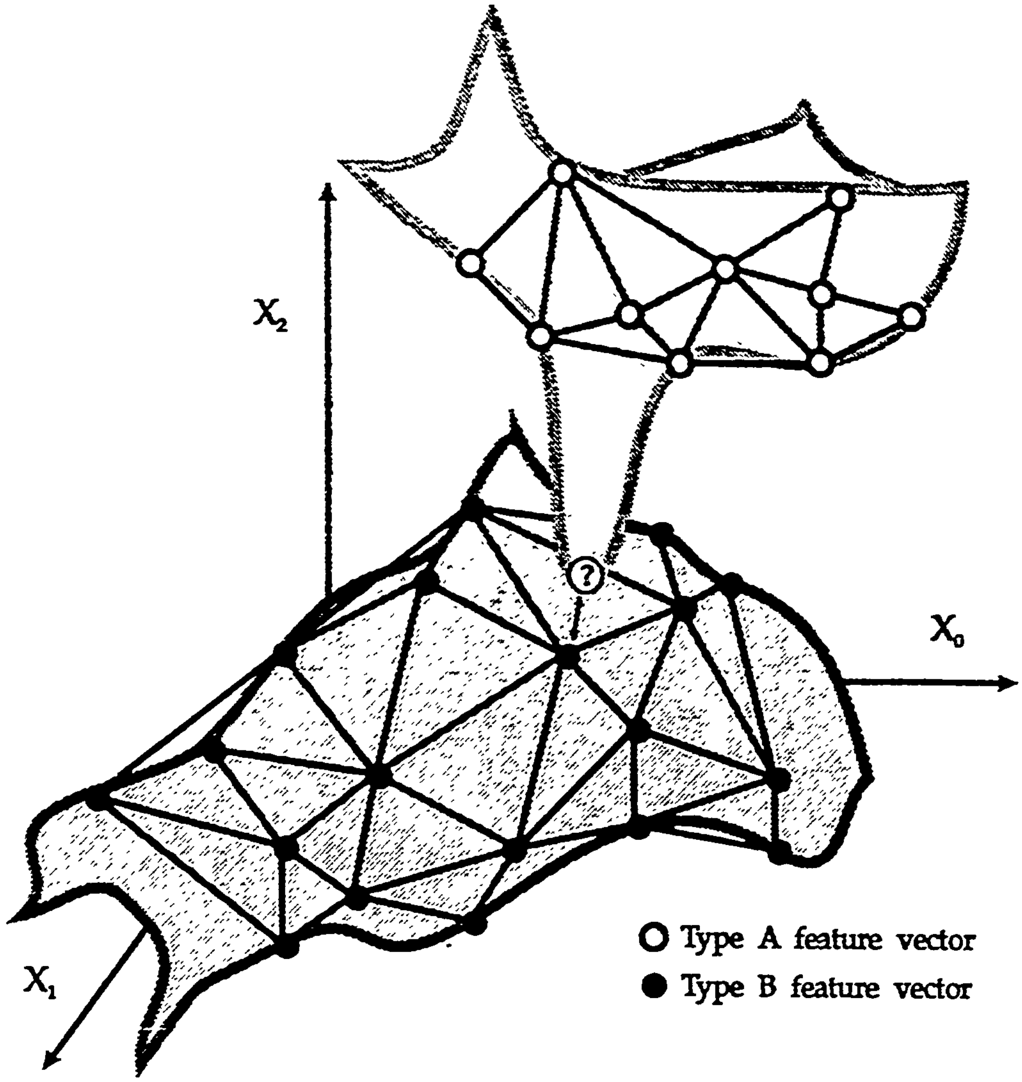


Figure 6. Confusion in feature space between two different objects A and B (an unknown feature vector of type A is incorrectly matched to a type B known feature vector because of an insufficient sampling of known feature vectors of type A. This is affected by both the topology of A's feature space and the proximity of b's surface.

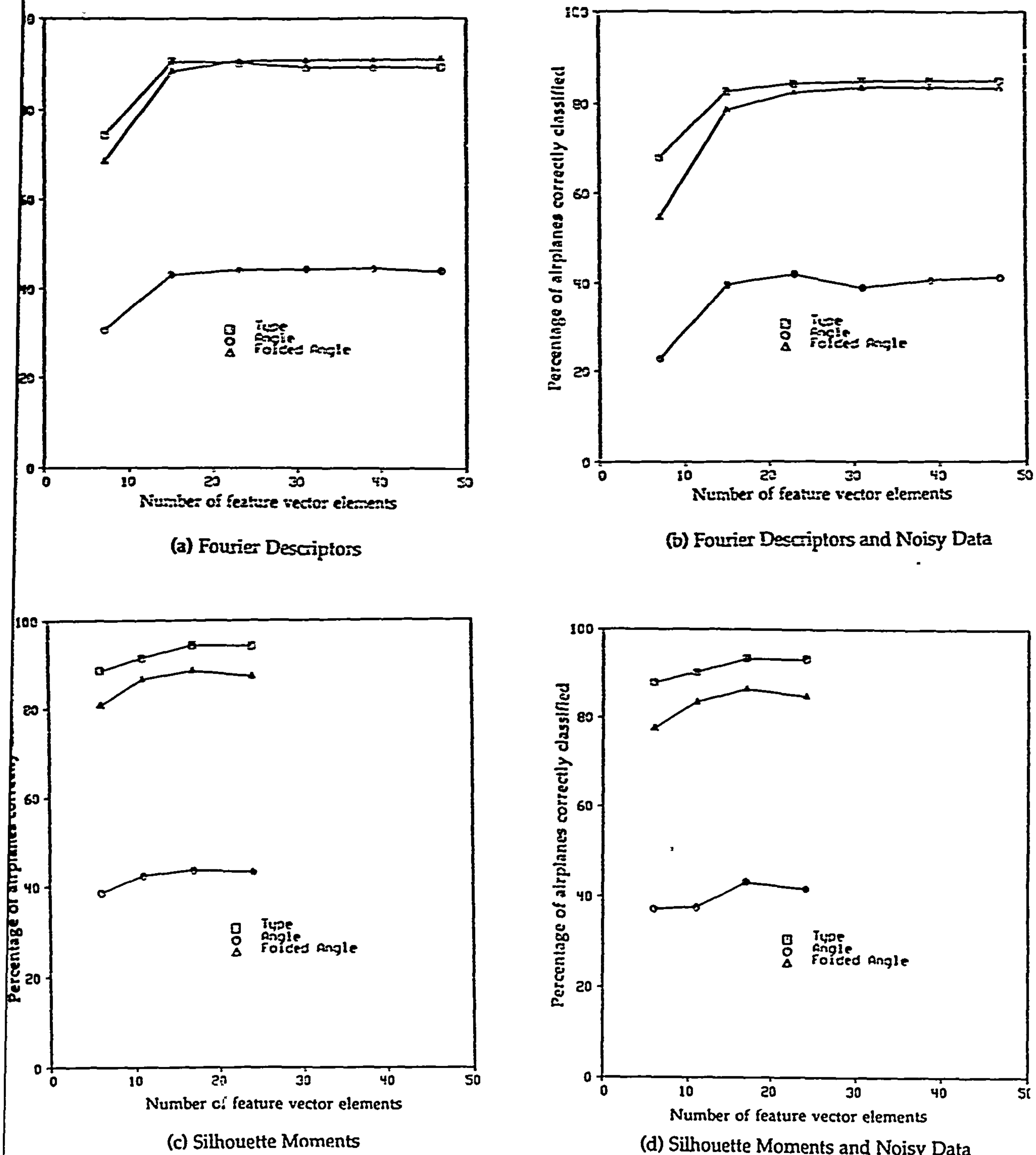
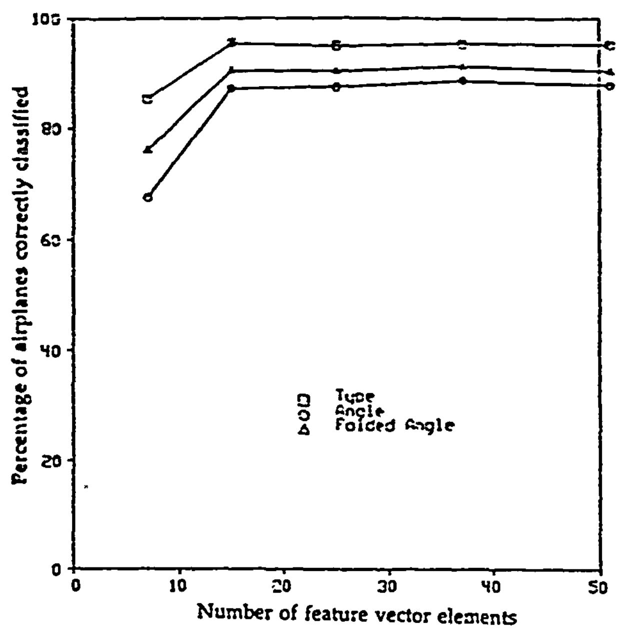
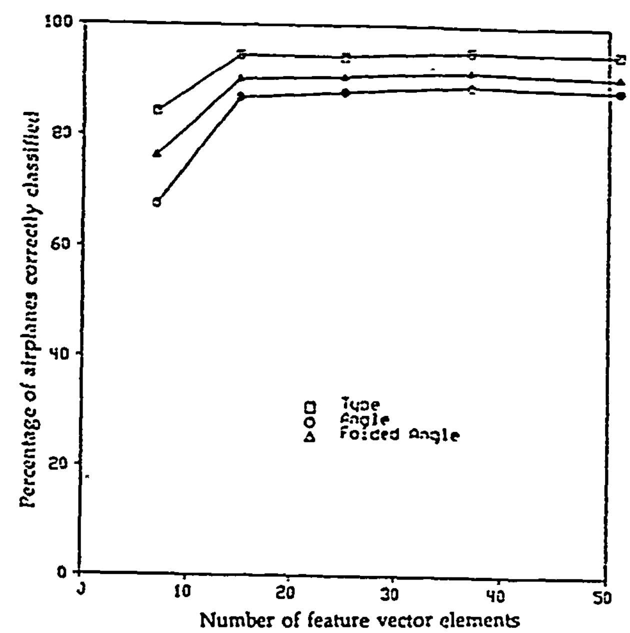


Figure 7. Classification results for the test data with and without added noise

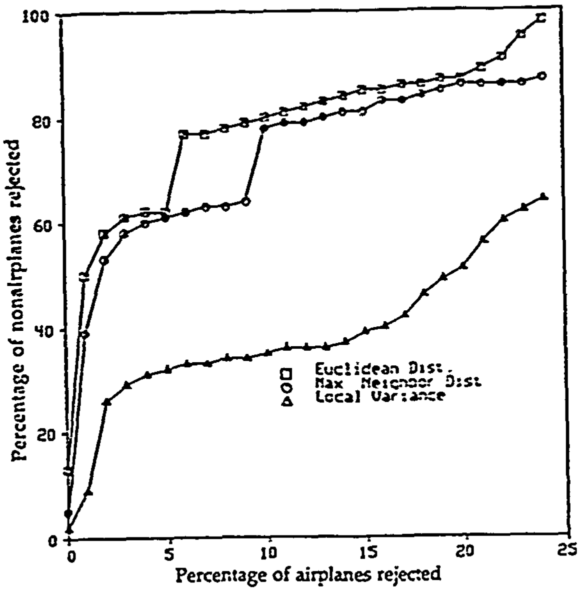


(e) Silhouette and Range Moments

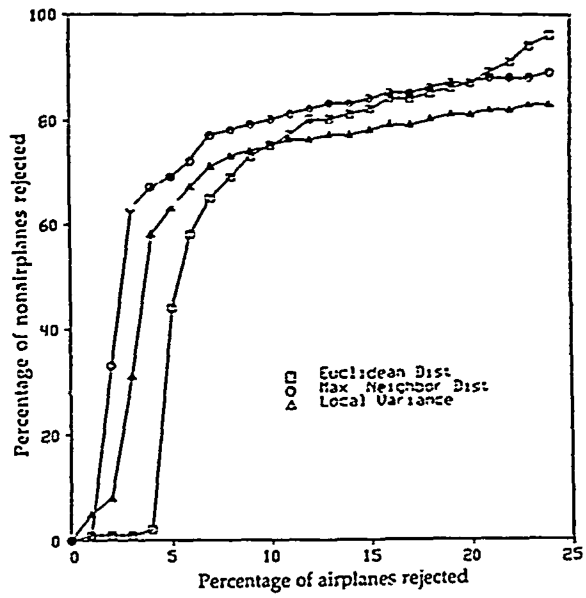


(f) Silhouette and Range Moments and Noisy Data

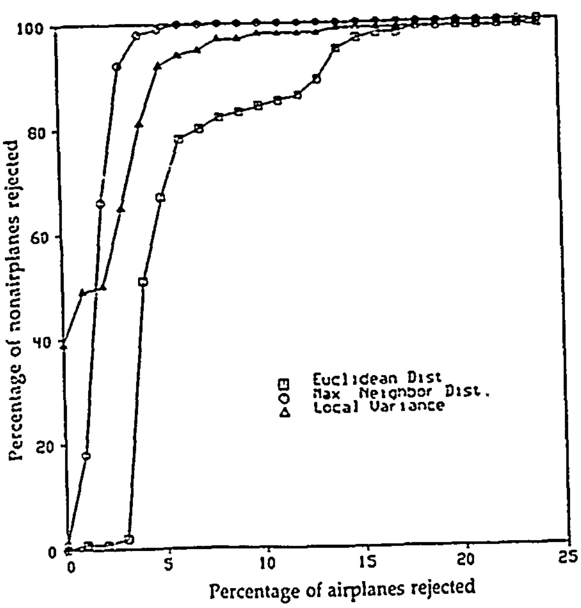
. Figure 7. continued



(a) Fourier Descriptors

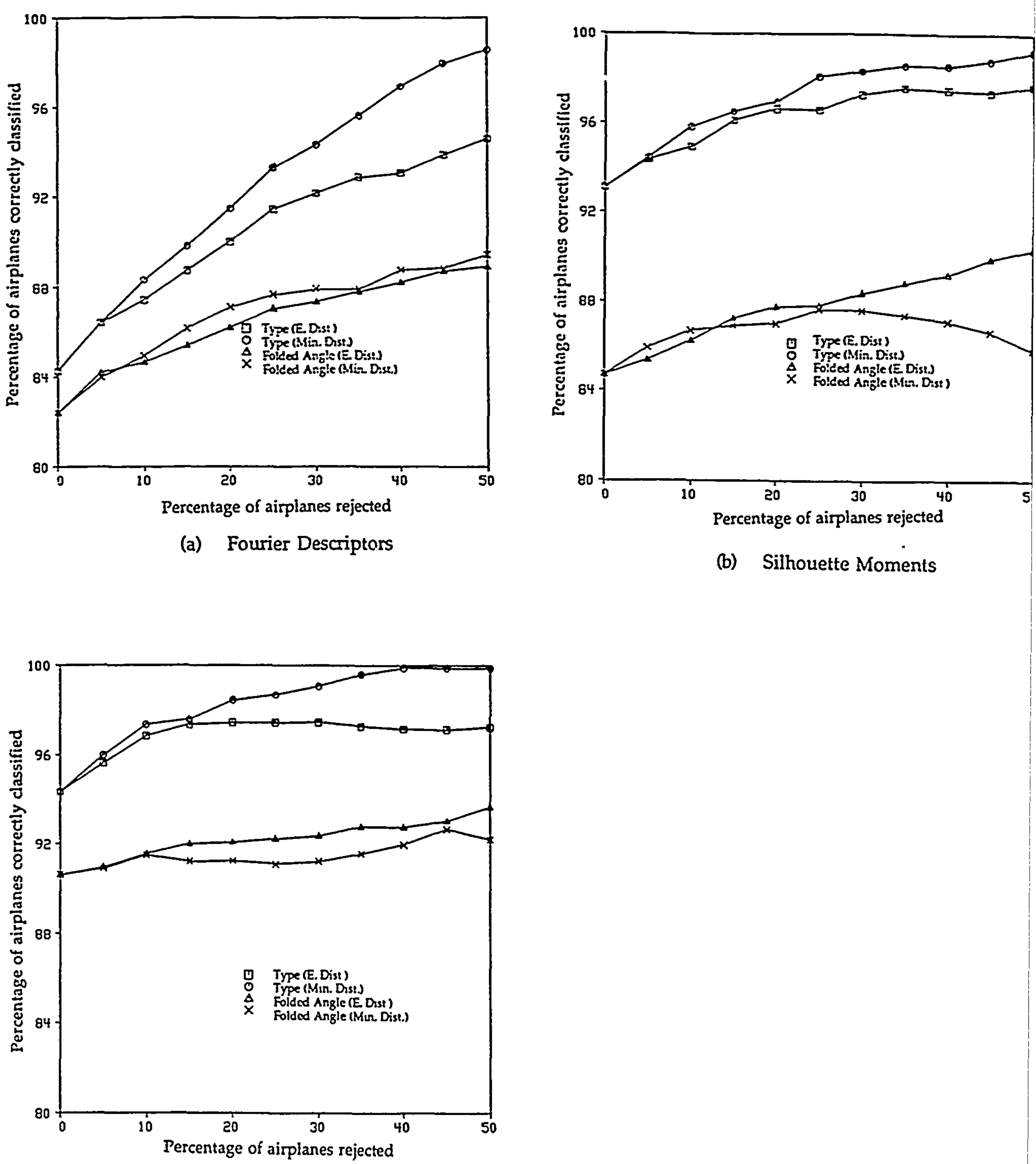


(b) Silhouette Moments



(c) Silhouette and Range Moments

Figure 8. Unknown object rejection results (CQA_1)



(a) Fourier Descriptors

(b) Silhouette Moments

(c) Silhouette and Range Moments

Figure 9. Effect rejecting low confidence responses on classification success (CQA)

REFERENCES

1. Reeves, A.P.; Prokop, R.J.; Andrews, S.E.; and F.P. Kuhl, " Three Dimensional Shape Analysis using Moments and Fourier Descriptors," Proceedings of the Seventh International Conference on Pattern Recognition, pp 447-450, July 1984.

2. Reeves, A.P., and Rostampour, A. " Shape Analysis of Segmented Objects Using Moments." 1981 Pattern Recognition and Image Processing Conference, pp 171-174, Dallas, Texas, 1981.

3. Dudani, S.A.; Breeding, K.J.; and McGhee, R.B."Aircraft Identification by Moment Invariants, " IEEE Transactions on computers **C-26**(1), pp 39-46 January 1977.

4. Wallace, T.P. and Wintz, P., "An Efficient, Three-Dimensional Aircraft Recognition Algorithm Using Normalized Fourier Descriptors," Computer Graphics and Image Processing **3**, pp 99-126, 1980.

5. Reeves A.P.and Taylor, R.W., "Identification of Three Dimensional Objects Using Range Information, " IEEE Transactions on Pattern Analysis and Machine Intelligence **11** (4), pp 403-410, 1989.

6. Kuhl, F.P.; Reeves, A.P.; and Taylor, R.W., Shape Indentification with Moments and Fourier Descriptors," Proceedings of the 1986 ACSM-ASPRS Convention, March 1986.

7. Taylor, R.W.; Reeves , A.P.; and Kuhl, F.P., " Classification Quality Assesss-ment for Object Identification Systems, " Proceedings of the 1987 ACSM-ASPRS Annual Convention **2**, pp 339-348, March 1987.

8. Taylor, R.W. and Reeves, A.P., " Generalizing Object Identification Systems Through Classification Quality Assessment, " IEEE Transactions on System Man Cybernetics **19** (4), pp 861-866, 1989.

9. Mitchell O.R.; Reeves, A.P.; and Grogan, T.A., "Algorithms and Architecture for Global Shape Analysis in Time-Varying Imagery," SPIE Proceedings: Robotics and Industrial Inspection **360**, pp 190-197, August 1982.

10. Ballard, D.H. and Brown, C.M., Computer Vision, Prentice Hall, 1982.

11. Yang, H.S. and Kak, A.C., " Determination of the Identity, Position and Orientation of the Topmost Object in a Pile, " Proceedings of the Third Workshop on Computer Vision: Representation and Control, pp 38-48,13-16 October 1985.

12. Ikeuchi,K. " Recognition of 3-D Objects Using the Extended Gaussian Image," Proceedings of the 7th IJCAI, pp 595-600, 24-28 August 1981.

13. Bhanu, B., " Representation and Shape Matching of 3-D Objects, " IEEE Trans. on Pattern Analysis and Machine Intelligence **PAMI-6** (3), pp 340-351, May 1984.

14. Jarvis, R.A., " A Perspective on Range Finding Techniques for Computer Vision," IEEE Trans. on Pattern Analysis and Machine Intellegence **PAMI-5** (2), pp 129-139 March 1983.

15. Yang, H.S.; Boyer, K.L.; and Kak, A.C.," Range Data Extraction and Interpretation by Structured Light," Proceedings of the First IEEE Conference on Artificial Intelligence Applications, pp 199-205, December 1984.

16. Flynn, P.J. and Jain, A.K.," On Reliable Curvature Estimation, " Proceedings of the IEEE Computer Society Conference on CVPR, pp 110-116, 1989.

17. Hu, M-K, " Visual Pattern Recognition by Moment Invariants," IRE Transactions on Information Theory **IT8**, pp 179-187, Febuary 1962.

18. Reeves, A.P. and Wittner, B.S.," Shape Analysis of Three Dimensional Objects Using the Method of Moments," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 20-26 June 1983.

19. Mitchelll, O.R. and Grogan T.A.," Global and Partial Shape Discrimination for Computer Vision," Optical Engineering **23**(5), pp 484-491, 1984.

20. Kuhl, F.P. and Giardina, C.P., " Elliptic Fourier Features of a Closed Contour, " Computer Graphics and Image Processing **18**, pp 236-258, 1982

Distribution List

Commander
Armament Research Developmentand Engineering Center
U.S. Army Armament ,Munitions and Chemical Command
ATTN: SMCAR-IMI (5)
SMCAR-FSA-SS (10)
SMCAR-CCH-P (2)
Picatinny Arsenal, NJ 07806-5000

Commander
U.S. Army Armament, Munitionsand Chemical Command
ATTN: AMSMC-GCL (D)
Picatinny Arsenal, NJ 07806-5000

Administrator
Defense Technical Information Center
ATTN: Accessions Division (2)
Cameron Station
Alexandria, VA 22304-6145

Director
U.S. Army Materiel Systems Analysis Activity
ATTN: AMXSY-MP
Aberdeen Proving Ground, MD 21005-5066

Commander
Chemical Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCCR-MSI
Aberdeen Proving Ground, MD 21010-5423

Commander
Chemical Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCCR-RSP-A
Aberdeen Proving Ground, MD 21010-5423

Director
Ballistics Research Laboratory
ATTN: AMXBR-OD-ST
Aberdeen Proving Ground, MD 21005-5066

Chief
Benet Weapons Laboratory, CCAC
Armament Research, Development and Engineering Center
U.S. Army Armament, Munitions and Chemical Command
ATTN: SMCAR-CCB-TL
Watervliet, NY 12189-5000

Commander
U.S. army Armament Munitions and Chemical Command
ATTN: SMCAR-ESP-L
Rock Island, IL 61299-6000

Director
U.S. Army TRADOC Systems Analysis Activity
ATTN: ATAA-SL
White Sands Missile Range, NM 88002